

Aufgabe 6: Transkribieren von Primärquellen mit TEI

Christof Schöch

DARIAH-DE

Lizenz: Creative Commons Attribution 4.0 (CC-BY).

Das vorliegende Dokument wurde für DARIAH-DE neu erstellt. Das vorliegende Dokument beruht teilweise auf den folgenden Materialien: James Cummings, Renée Baalen, Ylva Berglund-Prytz: "An Introduction to XML and the Text Encoding Initiative", Digital.Humanities@Oxford Summer School 2012, University of Oxford, 2012, <http://digital.humanities.ox.ac.uk/dhoxss/2012/>, CC-BY-SA.

Transkribieren von Primärquellen mit TEI

1. Lernziele

Nach dieser Übung sollten Sie in der Lage sein, die folgenden Dinge zu tun:

- * Die Konversion einer Transkription in eine TEI-Version vorzunehmen, bzw. eine einfache Transkription so anzureichern, dass sie den Schreibprozess mit abbildet.
- * Die Elemente `add`, `del` und `subst` sinnvoll einzusetzen
- * Die Elemente `choice` mit `abbr` und `expn` einzusetzen
- * Das Element `unclear` für schwer lesbare Stellen einzusetzen

2. Überblick

In dieser Übung werden Sie eine Transkription, für die lediglich ein Header in TEI vorliegt, in ein vollständig TEI P5-konformes Dokument verwandeln. Hierfür werden Sie die Elemente `add`, `del`, `subst`, `choice`, `abbr`, `expn`, und `unclear` in einfacher und verschachtelter Weise einsetzen, um auch etwas komplexere Phänomene kodieren zu können.

3. Los geht's!

Wir beginnen mit einer Beispieldatei, die ich für Sie vorbereitet habe. Es ist die Datei "`Preface-Start.xml`", die Sie bitte im Oxygen XML editor öffnen. Prüfen Sie nach, ob die Datei wohlgeformt und valide ist. Bevor wir loslegen, sollten wir ein Schema mit der Datei assoziieren, damit beim Kodieren nicht so leicht Fehler passieren. (Diesmal bekommen Sie ein fertiges Schema von uns!)

- * Wählen Sie in Oxygen im Menü: `Document > Validate > Validate with`.
- * Im Dialogfeld klicken Sie auf das Ordnersymbol und wählen die Datei "`tei_ms.rng`" aus, klicken auf `Open` und dann auf `OK`. Überprüfen Sie nun, ob das Dokument weiterhin valide ist (`Strg+Shift+V`).

Achten Sie auf Folgendes:

- * Es gibt einen minimalen Header, mit einer `msDesc` (mehr dazu später)

- * Es gibt eine auskommentierte "plain text"-Fassung der Edition dieses Vorworts, die von J. Stallworthy herausgegeben wurde. Die berücksichtigt nur den Text, den Stallworthy als letztlich intendiert verstanden hat.
- * Es gibt ein digitales Faksimile der Manuskript-Seite, das ist die Datei "Preface.jpg". Sie können sich diese Datei und die auskommentierte Textfassung einmal anschauen, und beide vergleichen. Welche Beziehung besteht zwischen den beiden?

4. Ein wenig Wiederholung: Textstruktur

Im Faksimile und in der Transkription kann man einen Titel und (mehr oder weniger eindeutig) mehrere Absätze erkennen. Stallworth unterscheidet fünf Absätze.

- * Entfernen Sie am Anfang und am Ende der Transkription die Zeichen für das "Auskommentieren" von Text: `<!-- ... -->`. Das Dokument ist jetzt nicht mehr valide.
- * Kodieren Sie den Titel und die Absätze mit den dafür gebräuchlichen Elementen von TEI: `head` für Überschriften, `p` für Absätze.
- * Die Überschrift ist unterstrichen; um solche Eigenschaften festzuhalten, verwendet man das Attribut `rend`, das Werte wie "italic" oder "underline" haben kann. Das Attribut weist darauf hin, wie eine Passage im Originaldokument dargestellt wurde; bei der Visualisierung kann man dies berücksichtigen oder nicht.
- * Das Element `head` sollte folgendermaßen aussehen: `<head rend="underline">Preface.</head>`
- * Überprüfen Sie, ob das Dokument immer noch valide ist. Wenn ja, formatieren Sie das Dokument neu (mit `Strg+Shift+P`) und speichern Sie das Dokument.

5. Zweiter Absatz: Löschungen und Ergänzungen

Der erste Absatz (bzw. Satz) hat keine Besonderheiten, die wir hier berücksichtigen müssen, der zweite Absatz ("Nor.." bis "...war.") aber schon. Man findet mehrere Streichungen und mehrere Ergänzungen. Die Streichungen sind nicht in der Transkription enthalten. Als Hilfestellung seien hier die gelöschten Passagen genannt: "battles, nor glory of battles, or land, or", außerdem "or" sowie "whatever".

- * Fügen Sie die Streichungen an der passenden Stelle des Textes ein
- * Kodieren Sie Streichungen und Ergänzungen nun mit `del` und `add`
- * Achten Sie immer darauf, dass ihr Dokument valide bleibt!

In zumindest einem Fall wird man davon ausgehen können, dass eine Streichung und eine Ergänzung zusammengehören, d.h. eine Ersetzung bilden.

- * Kodieren Sie dies, indem Sie die durch `add` und `del` kodierten Änderungen gemeinsam in ein Element `subst` einbetten.

6. Weitere Präzisierungen: Art der Änderungen

Mit `add` und `del` ist zwar die Funktion oder der Effekt einer Änderung kodiert, nicht aber, auf welche Art sie ausgeführt wurde. Dies sollten Sie in einem zweiten Durchgang durch den zweiten Absatz ergänzen.

- * Einige Ergänzungen stehen recht eindeutig höher als der übrige Text, andere tiefer. Geben Sie diesen Ergänzungen das Attribut `place`, mit dem Wert `above` bzw. `below` (weitere mögliche Werte könnten sind `left`, `right`, `etc.`; dies hängt vom Schema ab).

- * Das Element `del` bedeutet streng genommen nur "Löschung", sagt aber nichts über die Art der Löschung aus, die durch einfaches oder mehrfaches Durchstreichen, durch "Aus-X-en" oder andere Techniken vorgenommen werden kann. Fügen Sie dem Element `del` an den betreffenden Stellen das Attribut `rend` mit dem Wert `stroked` hinzu.

7. Weitere Präzisierungen: unleserliche Passagen

- * Einige der Wörter oder Passagen in der Handschrift sind nicht leicht zu entziffern. Zumindest an den problematischsten Stellen sollten Sie dies dokumentieren. Dafür gibt es das Element `unclear`.

- * Sie können außerdem auch den Grund angeben: bspw. weil die Tinte zu blass ist, oder weil die Handschrift schlicht unleserlich ist. Für einen solchen Grund gibt es das Attribut `reason`; geben Sie ihm einen Wert, der den Grund knapp benennt, bspw. "blass" oder "unleserlich". (Diese Werte entsprechen im Ernstfall einer von Ihnen erstellten Typologie, die Sie in einer Sprache Ihrer Wahl verfassen können, und die im Schema definiert und in der `encodingDesc` erläutert werden kann.)

8. Weitere Präzisierungen: Abkürzungen

- * In der vorliegenden Fassung des Textes wurde eine Abkürzung mit ihrer ausgeschriebenen Fassung transkribiert, und zwar im Falle des Wortes "dominion". Wenn man genau hinsieht, steht im Manuskript lediglich "domin"!

- * Kodieren Sie dies mit den Elementen `abbr` und `expn`. Da dies keine Änderung im Verlauf des Schreibprozesses ist, sondern zwei alternative Lesarten der Passage, sollten Sie beide Elemente in ein Element `choice` einbetten.

9. Zwischenbilanz und Stylesheet für Oxygen

Geschafft! Gehen Sie jetzt einmal von Oxygens "Text"-Modus in den "Autor"-Modus. Wie werden die Kodierungen hier angezeigt? Nicht sehr hübsch, oder? Laden Sie daher ein CSS-Stylesheet in Oxygen, mit dem die Transkription angemessen dargestellt wird. Gehen Sie wie folgt vor:

- * Kehren Sie zurück zum "Text"-Modus.
- * Wählen Sie "Document > XML Document > Associate XSLT/CSS Stylesheet"
- * Im Dateidialog wählen Sie den Reiter "CSS" aus.
- * Wählen Sie dann mit dem kleinen Ordnersymbol die Datei "tei-

transcr.css" aus, die sich ebenfalls in dem Ordner mit der Übung befinden. Klicken Sie auf `OK`, um zu bestätigen.

* Speichern Sie das Dokument, und wechseln Sie dann wieder in den `Author-Modus`.

Jetzt sollten die Bearbeitungen im Dokument besser sichtbar werden. Sie können hier im weiteren Verlauf Ihren Fortschritt beobachten. Schalten Sie nun aber wieder zurück zum `Text-Modus`.

10. Dritter Absatz

Im dritten Absatz ("Above all..." bis "...pity") findet man nun Löschungen und Ergänzungen, die teils ineinander verschachtelt sind. Das kommt vor, wenn jemand etwas hinzufügt, und sich dann anders entscheidet, und entweder etwas anderes hinzufügt, und/oder die Hinzufügung erneut durch etwas anderes ersetzt.

- * Fügen Sie die gelöschten Wörter an der passenden Stelle ein
- * Überlegen Sie sich, wie man diese Änderungen mit einer Kombination von `add`, `del` und `subst` angemessen wiedergeben könnte.
- * Kodieren Sie den Absatz.

11. Vierter Absatz (optional; sonst direkt zu Punkt 12)

Schon fertig!? Dann versuchen Sie sich doch an dem vierte Absatz ("Yet these" bzw. "I have" bis "truthful"). Als kleine Hilfe seien hier die gelöschten Passagen aufgelistet; achten Sie auf die Kombination der Elemente und die Position der Ergänzungen.

- * I have no hesitation in
- * making public / publishing such
- * My
- * past
- * not further consolation
- * past
- * The / This / a
- * bereaved
- * generation
- * If I thought the letter of this book would last, I woul might have used proper names;
- * to
- * children
- * War

12. Abschluss

Den letzten Absatz lassen wir einmal unbearbeitet. Vergleichen Sie nun erneut Ihre Transkription mit dem Faksimile. Was fällt Ihnen auf?

- * Bemerken Sie etwas, das noch kodiert werden sollte?
- * Hat die Transkription Fehler, die man korrigieren sollte?
- * Hat Wilfred Owen selbst bei den Überarbeitungen Fehler gemacht?

Vergleichen Sie Ihre Kodierung mit der Ihres Nachbarn / Ihrer Nachbarin.

- * Welche Unterschiede bemerken Sie?
- * Sind beide Kodierungen valide?
- * Entsprechen beide Kodierungen den Guidelines?

Das ist das Ende der Übung